ISSN 2186-7437

NII Shonan Meeting Report

No. 204

DevOps for Cyber-physical Systems

Sebastiano Panichella (Zurich University of Applied Sciences, Switzerland)
Paolo Arcaini (National Institute of Informatics, Japan)
Myra B. Cohen (Iowa State University, USA)
Aitor Arrieta (Mondragon University, Spain)

November 6–10, 2023

National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-Ku, Tokyo, Japan

Background and introduction

Context and goal

Cyber-Physical Systems (CPSs) combine digital cyber technologies (e.g., software, microprocessors, etc.) with physical processes (e.g., mechanical and electrical components). In many applications, CPSs interact with humans; thus making these CPSs safety-critical. Most of CPS functionality is driven by software, which requires maintenance and evolution to deal with hardware obsolescence, new legislation changes, correction of bugs, and inclusion of new functionalities. To better deal with maintenance and evolution, like in many other software domains, DevOps techniques have emerged. However, in the context of CPSs, these new software development techniques are still in their infancy, still requiring ground-breaking research solutions.

To this end, DevOps provides an effective solution for such continuous evolution. DevOps combines development practices and tools to speed up applications and service delivery while ensuring their high reliability. DevOps practices and tools found applications in different domains ranging from traditional systems to mobile applications. In the context of CPS, DevOps is also being considered as a natural choice to cost-effectively support CPS development and evolution. As a consequence, there has been an increasing interest in blending best practices from DevOps solutions with the development processes used in the CPS, to deliver software rapidly and dependably. However, doing so poses new challenges for both contemporary and future CPS development and operation that require developing novel research methods.

Existing DevOps solutions for CPSs suggest gaps that software engineering and CPS communities have to address in the future. Such challenges arise from the complexity and uncertainty of CPS behavior due to the CPS environment and the non-deterministic nature of such systems. Indeed, CPSs often rely on Artificial Intelligence (AI) based and sensor-based components, representing the main sources of non-determinism and uncertainty of CPSs within the operational environment. DevOps tools, with all their practices and expected automation, have a critical goal to enable the rapid and incremental delivery and assessment of CPS-developed software and hardware components, which are synergically used for enabling the effective operation of CPSs in their environment. At the same time, since most of the current behavioral assessment of such systems is based on manual and human-based activities, DevOps innovations have to also target automated ways on how to assess the behavior of CPSs with human-into-the-loop.

Given the relevance and multifaceted complexity of such challenges, it is required to determine the main aspects that need to be tackled by experts from different domains (DevOps experts, software engineering researchers, and CPS researchers/practitioners).

Details of the meeting

The core part of ongoing DevOps solutions requires advanced development and deployment strategies, which work efficiently and in an adaptive manner in different.

CPS domains while addressing different CPS technical requirements. Dur-

ing the development phase of CPS, it is common to apply different simulation techniques such as Software in the Loop (SiL) and Hardware in the Loop (HiL), e.g., for development and testing of CPS software. For example, HiL consists in establishing techniques used within the development and testing of complex real-time CPS by adding mathematical (or simulation model) representations of all related (physical) dynamics of such systems. Research in the field concerning DevOps automation for CPSs or simply running test cases in DevOps pipelines is still challenging, with development and deployment automation going beyond mere test execution to support HIL activities, which are now required for CPS.

This Shonan meeting provided a platform to discuss challenges and ideas in all aspects and phases of DevOps in the context of CPSs, including how DevOps can further contribute to improving the efficiency of tomorrow's CPS development and evolution. This facilitated the discussion along many dimensions of DevOps for CPSs, including DevOps pipelines (e.g., configuration management), functional and non-functional testing (including uncertainty testing), monitoring techniques, uncertainty detection, run-time verification, and maintenance. Moreover, we also discussed existing known techniques such as model-based testing, AI techniques (e.g., search algorithms and other machine learning algorithms), and digital twins.

With this Shonan Meeting, we focused on several follow-up activities. Among them, we focus on generating the first book on DevOps for CPSs. This book collects and integrates challenges, bad/best practices, experiences, tools, gaps, and future directions in the field identified during the meeting. The will be written in a coordinated way with participants and organizers of the Shonan meeting, and we expect it will guide future generations of experts moving forward on DevOps and testing automation for complex CPSs.

Our meeting attracted researchers from academia and industry who are experts in DevOps, CPS development, testing and operation, applied AI, and digital twins. The core team of the Shonan Meeting were researchers and industrial participants from the COSMOS H2020 project (`https://www.cosmos-devops.org/`) project, researchers and industrial practitioners from the Adeptness H2020 project (`https://adeptness.eu/`) project, and researchers on DevOps, CPS, AI, and DevOps for CPS from Asia, Europa, Australia, and the USA to cover different geographical areas. The list of participants include: Lei Ma, Zhenya Zhang, Shaukat Ali, Alessio Gambi, Annibale Panichella, Shane McIntosh, Vincenzo Riccio, Domenico Bianculli, Fabrizio Pastore, Aldeida Aleti, Birthe Böhm, Christian Birchler, John-Paul Ore, Antonio Cicchetti, Hironori Washizaki, Xiao-Yi Zhang, Stefan Klikovits, Thomas Laurent, Catia Trubiani, Xavier Devroey, Mattia Fazzini, Wesley Klewerton Guez Assunção, Sebastiano Panichella, Paolo Arcaini, Myra Cohen, Aitor Arrieta Marcos, Volker Stolz, Violet Ka I Pun, etc.

The meeting started with an introductory presentation of the organizers to set the expectations of the meeting. Then, each participant gave a talk on their research related to the meeting's topics, serving as a basis for discussion for the rest of the meeting. The meeting consisted of structured and open discussion sessions to exchange ideas in different settings, which served as the starting point for writing the planned book. We discussed the main topics to be used to design the outline of the book, with the idea to evolve it after the meeting together with potential contributors of chapters.

The concrete outcome of the meeting consists of a roadmap of follow-up activities, including the creation of educational material (i.e., the book mentioned

above and related educational artifacts), the organization of future international events and summer schools, and the writing of position papers and project proposals. From a broader point of view, we believe this meeting will provide a forum to build international cooperation and joint international projects and publications in the future between different communities.

	Agenda - Shonan meeting Nr. 204 - DevOps for Cyber-physical Systems - https://shonan.nii.ac.jp/seminars/204/		
	Day 1, Monday, November 6, 2023		
	Timeline	**Lead / Presenter(s)**	**Morning**
8:50	15-25 minutes	Shonan Organizers	Opening (Agenda - Focus of the Workshop - Book, Collaborations/Networking, etc.) → 15 mins
9:15	1h 30 minutes	Attendees	Individual Presentations on Topics concerning DevOps automation in relations with CPS challenges 5 presentations
10:45	30 minutes		**break**
11:15	1h 30 minutes	Attendees	Individual Presentations on Topics concerning DevOps automation in relations with CPS challenges 5 presentations
12:30	1h		**Lunch break**
	Timeline	**Who**	**Afternoon - Discussion on possible topics**
14:15	1h 30 minutes	Attendees	**Individual Presentations on Topics concerning DevOps automation in relations with CPS challenges 5 presentations**
15:45	30 minutes		**break**
16:30		All (different groups)	**Individual Presentations on Topics concerning DevOps automation in relations with CPS challenges 5 presentations**
17:45		All (different groups)	**(Participants) break-out group discussion / (Organizers) Book/report topics: (- Datasets and benchmarks in different CPS domains, - DevOps challenges/stages, - etc.)**
	Day 2, Tuesday, November 7, 2023		
	Timeline	**Lead / Presenter(s)**	**Morning**
9:00	1h 45 minutes	Attendees	Individual presentations (5 presentations)
10:45	30 minutes		**break**
11:25	1h 10 minutes	All	Individual presentations (4 presentations)
12:30	1h		**Lunch break**
	Timeline	**Who**	**Afternoon - Discussion on possible topics**
13:50	1h 55 minutes	Attendees/all	Individual presentations (2 presentations) Finalization of the topics for breakout discussion and formation of the groups
15:45	30 minutes		**break**
16:15	2h	All (different groups)	**(Participants) break-out group discussion / (Organizers) Book/report topics: (- Datasets and benchmarks in different CPS domains, - DevOps challenges/stages, - etc.)**
18:00	1h 30 minutes		**Dinner**
	Day 3, Wednesday, November 8, 2023		
	Timeline	**Lead / Presenter(s)**	**Morning**
9:00	1h 45 minutes	All (different groups)	(Participants) break-out group discussion / (Organizers) Book/report idea
10:45	30 minutes		**break?**
11:15	1h 30 minutes	Attendees	Individual presentations
11:30			**Group pictures of participants!**
12:30	1h		**Lunch break**
		All	Excursion "Visiting the Great Buddha and Hase Temple"
	Day 4, Thursday, November 9, 2023		
	Timeline	**Lead / Presenter(s)**	**Morning**
9:00	1h 45 minutes	Attendees	Individual presentations
10:45	30 minutes		**break**
11:15	1h 30 minutes	Attendees	Individual presentations
12:30	1h		**Lunch break**
	Timeline	**Who**	**Afternoon - Discussion on possible topics**
14:00		All (different groups)	(Participants) Finish break-out group discussion / (Organizers) Book/report idea
15:30	30 minutes		**break**
16:00		All (different groups)	(Participants) Start last break-out group discussions / (Organizers) Book/report idea
	Day 5, Friday, November 10, 2023		
	Timeline	**Lead / Presenter(s)**	**Morning**
9:00	1h 45 minutes	All	Group presentations / Individual presentations
10:45	30 minutes		**break**
11:15	1h 30 minutes	All	Other - Other collaboration and outcomes - Open discussion / Feedback - Closing
12:30	1h		**Lunch before leaving the venue?**

Figure 1: Overview of schedule of the meeting

Overview of the meeting and schedule

The meeting started with an introductory presentation of the organizers to set the expectations of the meeting. Then, each participant gave a talk on their research related to the meeting's topics, serving as a basis for discussion for the rest of the meeting. The meeting consisted of structured and open discussion sessions to exchange ideas in different settings, which served as the starting point for writing the planned book. We discussed the main topics to be used to design the outline of the book, with the idea to evolve it after the meeting together with potential contributors of chapters. In Figure 1 the structure of the Shonan meeting 204.

Overview of Talks

Testing AI-based cyber-physical systems

Aldeida Aleti, Monash University

Social robots, defined as cyber-physical systems capable of autonomous interaction with human or non-human agents in real environments, are finding primary applications in nursing homes, hospitals, and private residences. These applications aim to provide crucial assistance to diverse populations, including the elderly, individuals with disabilities, children, and medical personnel.

Despite the potential rapid integration of social robots into environments critical to the well-being of individuals, concerns about their reliability persist. Given the nascent nature of this market, there is an urgent need for comprehensive research and development to ensure the safety and reliability of these robotic assistants in various settings. The potential for technical malfunctions, misinterpretation of cues, or unintended consequences in dynamic and unpredictable real-world scenarios highlights the need for rigorous testing and validation. Reliability concerns extend beyond mere technical functionality; they encompass the ability of social robots to consistently deliver assistance, companionship, and care without compromising the well-being of users. Imagine a social robot deployed in a care facility to assist elderly residents with daily tasks and provide companionship. During the testing phase, the robot successfully navigates through predefined scenarios and engages in scripted interactions. However, in a real-world scenario, the environment is dynamic, and the needs of the elderly residents can be unpredictable. A technical malfunction, such as a sensor failure or a software bug, could result in the robot misinterpreting its surroundings. Misinterpretation of cues becomes a significant concern, especially when the robot is designed to recognize and respond to human emotions. If the robot misinterprets a resident's emotional state, it might provide inappropriate responses or fail to offer the needed support. The robot uses natural language processing to understand and generate responses. However, during the development phase, the training data used to teach the robot's language model inadvertently reflects biases present in the data. The robot's emotional recognition system might be biased in its interpretation of facial expressions or tone of voice. Unintended consequences may also arise in scenarios where the robot attempts to assist with physical tasks. For example, if the robot is programmed to help a resident stand up, a miscalculation or misjudgment in force application could result in discomfort or even injury. To address these challenges, the aim of this talk is to discuss new testing and validation approaches for cyber physical systems such as social robots which generate a wide range of realistic scenarios, including unexpected and safety critical situations, to ensure the robot's robust and reliable performance in dynamic environments.

Uncertainty Unveiled: Identifying and Classifying Industrial Cyber-Physical Systems' Uncertainties for Testing

Shaukat Ali, Simula Research Laboratory and Oslo Metropolitan University

Cyber-physical systems (CPSs) are part of our daily lives. Such CPSs are prone to various uncertainties (e.g., uncertainties in hardware and operating

environment) that can affect their performance and safety in the worst case. Thus, it is essential to identify and understand various uncertainties for the CPSs during their design and operation time and use them to derive tests. Such tests can ensure that the CPSs handle uncertainties to avoid failures and ensure optimal performance. This presentation focuses on a step-wise process of testing CPSs (in particular in the vertical transportation domain) in Software in the Loop (SiL) configuration, mainly focusing on (1) the identification of various uncertainties, (2) classifying uncertainties according to various degrees of risks, and (3) generating uncertainty-aware tests to test CPSs under these uncertainties.

Repair Techniques for Deep Neural Networks for Safety in Perception of Driving Systems

Paolo Arcaini, National Institute of Informatics

Deep Neural Networks (DNNs) are often used in safety-critical systems, such as autonomous driving. Hazards of these systems are usually linked to specific error patterns of the DNN, such as specific misclassifications. In the context of the eAI project–"Engineerable AI Techniques for Practical Applications of High-Quality Machine Learning-based Systems" (`https://www.jst.go.jp/mirai/en/program/super-smart/JPMJMI20B8.html`), we are investigating techniques to repair a DNN to fix some given misclassifications that are considered particularly critical by stakeholders. The first step of these repair approaches consists in applying fault localization (FL) to identify the DNN components (weights) responsible for the misclassifications. However, the components responsible for one type of misclassification could be different from those responsible for another type; depending on the granularity of the analyzed dataset, FL may not reveal these differences: failure types more frequent in the dataset may mask less frequent ones. The talk will first present a way to perform FL for DNNs that avoids this masking effect by selecting test data in a granular way. Then, the talk will introduce DistrRep, a distributed repair approach that considers the fact that different misclassifications lead to different hazards, and so are associated with different risk levels. DistrRep first finds the best fixes for each critical misclassification, and then integrates them in a single repaired DNN model, by considering the risk levels. Finally, the talk will discuss a problem that affects DNN repair: as the search progresses and the network evolves, the weights responsible for the faults in the system will change, and the search will lose in effectiveness. The talk will introduce AdRep, an adaptive search method for DNN repair that adaptively updates the target weights during the search by performing fault localisation on the current state of the model.

DevOps for Cyber-Physical Systems: Results and Lessons Learned from the Adeptness H2020 Project, and the road ahead

Aitor Arrieta, Mondragon University

While most large web-based software systems (e.g., Amazon, Google) release a new software version every almost a minute, in the context of Cyber-Physical

Systems (CPSs), this is still far. However, the software of CPSs needs to evolve while these are in operation to fix bugs, add new functionalities, carry out refactoring activities and deal with unforeseen situations that were discovered while the CPS was operating. In the last three years, the Adeptness project has been developing a solution to help speed-up the software release of CPSs that are in operation while guaranteeing their reliability. In this talk, we summarize the objectives, results and lessons learned from this H2020 project, and we explain the road and challenges ahead.

Traceability: is everything connected (or should it be)?

Wesley Assunção, North Carolina State University

Traceability has already been proven as paramount for the development of traditional software, helping the understanding, validation, testing, and maintenance of systems. However, creating and maintaining trace links among artifacts is acknowledged as a challenge, because of the diversity of tools, stakeholders, and artifact types. In the context of Cyber Physical Systems (CPSs), traceability is more challenging, as by default CPSs are multidisciplinary, not only being composed by software, but also having pieces of hardware, sensor, actuators. Yet, the development of CPSs comprehends the collaboration among such disciplines, requiring the traceability among software artifacts, electric and mechanic models, complex computation tools, and the like. Thus, new challenges emerge, which must be discussed and addressed. This talk presents some past work on traceability, describes some existing challenges in the state of the art and the practice, and highlights the need of proper traceability solutions in the context of DevOps for CPSs.

Signal-based requirement properties for CPS: specification, monitoring, and diagnostics

Domenico Bianculli, University of Luxembourg

Run-time verification (RV) is an analysis technique that focuses on observing the execution of a system to check its expected behavior against some specification. It is used for software verification and validation activities, such as operationalizing test oracles and defining run-time monitors.

The three main components of an effective RV approach are: i) a specification language allowing users to formally express the system requirements to be checked; ii) a monitoring algorithm that checks a system execution trace against the property specifications and yields a verdict indicating whether the input traces satisfies the property being checked; iii) a diagnostics algorithm that explains the cause of a requirement violation, in case of a negative verdict.

In this short talk, I will review these three aspects taking into account the perspective of signal-based requirement properties in the CPS domain, and discuss how they can fit within DevOps practices.

Regression Testing for Self-driving Cars in Simulation Environment

Christian Birchler, Zurich University of Applied Sciences

Simulation platforms facilitate the development of emerging cyber-physical systems (CPS) like self-driving cars (SDC) because they are more efficient and less dangerous than field operational test cases. Despite this, thoroughly testing SDCs in simulated environments remains challenging because SDCs must be tested in many long-running test cases. Past results on software testing optimization have shown that not all test cases contribute equally to establishing confidence in test subjects' quality and reliability, and the execution of "safe and uninformative" test cases can be skipped to reduce testing effort. However, this problem is only partially addressed in the context of SDC simulation platforms.

Environmental sustainability and DevOps for cyber-physical systems

Birthe Böhm, Siemens AG

Environmental sustainability gets more and more important for our future, and a lot of companies are addressing this issue, including Siemens. Making software more environmentally friendly and climate-conscious means reflecting the whole lifecycle with respect to its carbon emissions and, inherently, its energy consumption. Therefore, also DevOps is expected to contribute – but which are the most promising approaches? That's why we started with researching on sustainable software engineering.

Software products play a considerable role in both global electricity demand and carbon emissions and, therefore, can contribute to these targets. Making software products more environmentally sustainable means reflecting their whole lifecycle with respect to their carbon emissions and, inherently, their energy consumption. There are already some standards available or in definition which address lifecycle assessments of software products, such as the Software Carbon Intensity Specification by the Green Software Foundation, or the Greenhouse Gas Protocol's ICT Sector Guidance published by the Carbon Trust and Global e-Sustainability Initiative. However, measuring is just one side of the medal. In order to reduce carbon emissions of software products, environmental sustainability must be established as a built-in quality in software engineering from product planning to operations.

Our goal is to set up a sustainable software engineering framework that guides organizations in creating sustainable software products. DevOps is essential for software engineering and, therefore, can contribute. We have collected ideas such as change-based pipeline adaptations, change-based testing, and energy-consumption-aware configuration of resources which are based on existing best practices. This talk wants to put environmental sustainability up for discussion in DevOps for cyber-physical systems to jointly fight global warming.

A reality check on DevOps adoption in industry

Antonio Cicchetti, Mälardalen University

The state-of-the-art in Software Engineering shows already a large corpus of research results on the benefits of adopting DevOps, even for CPS development. However, when trying to transfer these practices in industry the situation gets, by far, more complex than expected. In this short presentation I will discuss several industrial settings we have faced and are facing with our partners and will highlight the critical role played by consistency management to enable DevOps adoption in practice. Moreover, I will describe how AIDOaRt, a large ongoing EU project, is trying to address these issues with the support of AI/ML.

Testing Cyber-Physical Systems: Choose wisely

Xavier Devroey, University in Namur

The impacts of unexpected behavior from Cyber-Physical Systems (CPSs) can lead to significant damage like disruption of a production line, overheating of a nuclear reactor, false fire alarm, etc. That is why the safety and security of such systems should be at the center of concerns. Given the highly-configurable nature of CPSs, testing such systems is not trivial. Fuzzing works particularly well with any system by sending pseudo-random inputs. To adapt to specific systems and test requirements (coverage, resources, etc.), fuzzing is itself highly-configurable (Grammar-based, symbolic, probabilistic, etc.). This is why it could perform particularly well with CPSs, which all might require a different and specific testing approach depending on their interfaces, components, etc. In our work, we seek to define a configurable approach to find and recommend the most suitable fuzzing techniques for a given CPS and a set of test requirements.

Automated Corrective Maintenance

Mattia Fazzini, University of Minnesota

Because exhaustive testing is not generally possible, software is released with bugs and these bugs will translate into field failures. The ability to react effectively to field failures is, therefore, also essential to resolve bugs, but the support for this task is still limited and based on mostly manual, human-intensive approaches.

In this presentation, I will present two techniques for improving corrective maintenance tasks. The first technique automatically translates natural-language bug reports into test cases, so that developers can use the generated tests to focus their attention on debugging failures and quickly fix their apps. The second technique aims to improve the quality of submitted bug reports by providing suggestions as reporters type the bug reports. The techniques target mobile apps and the presentation makes technical connections on how to extend the approaches to cyber-physical systems.

On the Automated Generation of Critical Scenarios for Testing Self-Driving Car Software

Alessio Gambi, IMC University of Applied Sciences Krems

Critical scenarios can be used to test AVs in situations that are difficult to observe and are too expensive/dangerous to reproduce in real life. Because data collected by sensors are generally unavailable, we need to devise techniques that derive, mine, or extract critical scenarios from other sources, for example police reports, and optimize them towards specific testing goals. In this work, I'll summarize my ongoing research on reconstructing critical scenarios from text and sketches and their optimization using search algorithms.

Harnessing the power of Digital Twins

Stefan Klikovits, Johannes Kepler University Linz

Recent years saw the development and use of Digital Twins in many cyber-physical systems. Thereby, it is important to remember that the current view expands on the view that DTs are "convenient interfaces" to simulations, but enable the exploitation of powerful DT services that include runtime monitoring, planning, testing, log analysis, configuration, calibration, virtual commissioning, etc. Being a fundamental service of (most) DTs, simulation thereby serves as enabler for many services, but can be seen as "just one tool" in the DT toolbox, similar to AI learning.

In my research, I use DTs as bridging technology that enables CPS builders to use these powerful services, allow them to compose individual DT enhanced components to DT systems and systems of systems. I believe this work relates closely to the field of DevOps that heavily relies on the availability of the above services.

Mutation analysis for autonomous driving path planning

Thomas Laurent, JSPS fellow

Autonomous Driving Systems (ADS) are being deployed on the road more and more, and several accidents have already shown the importance of thoroughly testing them. One of the central components of these systems that can be tested early in development through simulation is the path planning component, which decides the car's target trajectory. As driving scenarios, i.e., tests are very complex in ADS simulation-based testing, it is very complex to know when one has sufficiently tested their system, or what gaps a test suite does not cover. Code-based mutation analysis has proven to be a strong test criterion in many domains, but is ill suited to the complex and peculiar structure of most path planning modules. We proposed a way of adapting mutation analysis to the path planner testing problem by mutating system parameters that reflect the system's decision making process. This allows us to assess if a test suite correctly covers the system's different behaviours. We showed that this criterion could both give feedback on what behaviours a test suite does (not) cover, and that it could be used to drive an automated test generation process using two industry-grade path planning systems and simulators.

The Road towards Trustworthy AI-enabled Cyber-Physical Systems

Lei Ma, The University of Tokyo / University of Alberta

A brief summary on our past and ongoing efforts to design and develop trustworthiness assurance methodology, techniques, and toolchain to provide more systematic support for the AI-CPS lifecycle.

(Getting Back to) Incremental Builds for DevOps Pipelines

Shane McIntosh, University of Waterloo

The incremental build was a key feature of build automation tools in the past. It still plays a key role in the build systems that underpin CI/CD and DevOps pipelines. Yet it is quite common for these "upper layer" automation technologies to start from a clean copy of the codebase, rendering the incremental features of build tools inert. In this talk, I will describe past and ongoing work from my group that strives to make CI/CD pipelines operate incrementally again.

Towards Adages for Cyber-Physical DevOps

John-Paul Ore, North Carolina State University

There's magic in the interplay of software and hardware. This talk speculates on a future when DevOps for Cyber-Physical Systems has solidified enough so that adages (sayings that consolidates wisdom) can describe the core concepts of DevOps for CPS. Additionally, this talk summarizes some of my current research and shows some entertaining videos of my flying robots crashing catastrophically into lakes.

Handling the many levels of testing for CPS

Annibale Panichella, Delft University of Technology

Testing for Cyber-Physical Systems (CPS) presents critical challenges, primarily related to the high costs, notably in terms of running time and resources. This presentation focuses on addressing these challenges through two distinct yet complementary approaches. The first approach tackles the issue by employing test decomposition, which involves extracting numerous fast unit-level tests from a limited number of resource-intensive system-level tests. Our findings suggest that these carved tests effectively complement manual and automatically generated unit tests (with fuzzing methods) and are preferred by developers for their improved readability. The second strategy, multi-level regression testing, aims to prioritize test cases of different granularity levels using diversity-based black-box strategies. To bridge the vocabulary gap between tests of varying levels, we leverage the WordNet ontology. Our initial analysis indicates that incorporating WordNet improve tests ordering, leading to the earlier discovery of regression faults compared to conventional approaches based on textual similarity.

Development Challenges for Complex Cyber-Physical Systems: Insights from the COSMOS H2020 Project

Sebastiano Panichella, Zurich University of Applied Sciences

Over the past decade, the development of Cyber-Physical systems (CPSs) has enabled significant advancements in healthcare, avionics, automotive, railway, and robotics. Notably, Unmanned Aerial Vehicles (UAVs) and Self-driving Cars (SDCs) have emerged as the frontrunners in avionics and automotive sectors, showcasing autonomous capabilities through onboard cameras and sensors. These systems have opened doors to a range of applications, including crop monitoring, medical and food delivery, and 3D reconstruction of archaeological and space exploration sites. However, state-of-the-art technology still lacks solutions that can operate in real-life missions due to limited testing solutions, which remains the biggest challenge.

This presentation will discuss the testing and development challenges faced by the COSMOS H2020 Project (`https://www.cosmos-devops.org/`) in the context of complex CPSs. COSMOS brings together a consortium of four academic and eight industrial partners, with organizations from the healthcare, avionics, automotive, utility, and railway sectors. The talk will focus on the studies conducted by COSMOS to identify the types of bugs affecting CPSs and the safety-critical issues of UAVs, along with automated testing approaches for UAVs, addressing the issue of the "Reality-gap".

Mutation Analysis and Testing for CPS

Fabrizio Pastore, University of Luxembourg

Software for Cyber-Physical Systems (CPS) is largely verified and validated through testing; satellite software systems are a typical example. Assessing the quality of test suites is therefore of crucial importance not only for developers but also for regulatory agencies such as the European Space Agency.

Known solutions to compensate for the limitations of current coverage-based practices are mutation analysis, which consists of measuring the mutation score (i.e., the proportion of injected faults detected by a test suite), and mutation testing, which consists of automatically generating test cases to maximize the mutation score. However, existing techniques can't address large and complex CPS; for example, mutation analysis may not terminate in a reasonable time.

Recent solutions to address the problems above range from relying on the fixed size confidence interval approach to quickly estimate the mutation score, mutating data to emulate interoperability problems, and relying on evolutionary algorithms (fuzz testing) to generate test cases.

Automated Workflow Planning for DevOps

Violet Ka I Pun, Western Norway University of Applied Sciences

A DevOps pipeline can be seen as a workflow consisting of a set of automated processes and tools such that developers and operation professionals can collaborate to build and deploy program code to a production environment.

Formal techniques has recently been applied for workflow modelling by using formal abstract executable models and leveraging concurrency theory analyses. In this talk, I would like to discuss how DevOps can be captured in the form of workflows using formal methods.

Driving Towards Quality: Evaluating Autonomous Vehicle Performance through Adequate and Diverse Tests

Vincenzo Riccio, Università di Udine

Among the huge variety of existing safety-critical Cyber-Physical Systems (CPSs), autonomous vehicles are steadily growing in relevance within the Software Engineering research community. Researchers proposed several test generation approaches to assess the quality of autonomous driving agents. Tool competitions, such as the one hosted by the SBFT workshop, offer the opportunity to benchmark new testing approaches and improve the state of the art. In this talk, I will focus on the problems of (1) assessing the diversity of the generated test cases and (2) evaluating the adequacy of a test suite when testing with autonomous vehicles. Hopefully, by addressing these open problems, our research community will propose more efficient and effective test generators and, as a consequence, improve the quality of driving agents.

Hardware-assisted Monitoring and Runtime Verification

Volker Stolz, Western Norway University of Applied Sciences

This talk gives a short overview on applying formal modelling to hardware- and software design. In particular, I illustrate two aspects related to DevOps and/or CPS: automated refactorings and runtime verification for low-level processor-trace mechanisms on SoCs. Especially for the latter, there's the interesting question on how monitors are designed, managed and deployed as part of a DevOps pipeline.

Performance Modeling of Software Performance Antipatterns for Cyber-Physical Systems

Catia Trubiani, Gran Sasso Science Institute

Interpreting the performance characteristics of software systems is not trivial, even more so when looking at specific application domains, such as cyber-physical systems whose complexity is exacerbated by the heterogeneity of software and hardware components and uncertainty in the operational environment. Our research focuses on developing methods to explain the performance analysis results and relate them with the system design alternatives. One viable solution is to make use of software performance antipatterns since they are recurring solutions to common mistakes, i.e. bad practices, affecting performance. Such antipatterns can play a key role in the software performance domain, because they can be used in the detection of performance issues as well as in the formulation of solutions, i.e., different architectural options. Recent work includes the performance modelling and analysis of eight software performance antipatterns,

along with a sensitivity analysis that investigates the peculiar characteristics of antipatterns, e.g., the frequency of checking the status of resources. This way, quantitative information to software designers can be provided to identify potential performance problems and their root causes.

SWEBOK evolution and Integrated platform for multi-view system modeling and machine learning pipelines

Hironori Washizaki, Waseda University

This talk firstly overviews the SWEBOK guide and its latest updates, including emerging topic areas, including DevOps and IoT. Furthermore, the talk presents the Multi-view Modeling Framework for Machine Learning Systems (M3S) with the ML pipeline integration to address the probabilistic nature of machine learning (ML) and its experimentative development approach, often resulting in a disparity between the quality of ML models with other aspects such as business, safety, and the overall system architecture. M3S provides an analysis framework that integrates different views. It is supported by an integrated metamodel to ensure the connection and consistency between different models. To facilitate the experimentative nature of ML training, M3S provides an integrated platform between the modeling environment and the ML training, performance monitoring, and repair pipelines.

Towards Understanding the Potential Failures of Complex Intelligent Systems

Xiao-Yi Zhang, University of Science and Technology Beijing

In the era of IoT 4.0, software becomes more intelligent and roles with the potential for significant impact. However, failures in such software products may lead to destructive loss. Yet, the complexity of current software systems, like Cyber-Physical Systems (CPS), makes addressing system failures a significant challenge. Specifically, conducting effective testing to reflect real-life failures can be difficult. Moreover, localizing and explaining the root cause of failures can be equally challenging. In this context, I would like to share some of our efforts in addressing failures within complex intelligent systems. These efforts encompass research works such as testing autonomous driving systems (ADS), analyzing reinforcement learning agents in complex environments like the Go game, and localizing components within ADS and deep neural networks (DNNs). Furthermore, I will introduce our thoughts on future initiatives and research directions aimed at testing and analyzing complex intelligent systems toward a better understanding of their failures.

Falsification of AI-Enabled Hybrid Systems

Zhenya Zhang, Kyushu University

Modern Cyber-Physical Systems (CPSs) (e.g., autonomous driving) are increasingly using AI-enabled controllers, mainly based on deep neural networks (DNNs). Falsification is an established approach for CPS quality assurance;

however, it is often guided by robustness derived from specification and only based on system output, and the temporal internal behavior of neural network controllers is not explored sufficiently. To bridge this gap, we make an attempt at exploring the temporal behavior determined by the repeated executions of the neural network controllers in hybrid control systems and propose time-aware coverage criteria for neural network controllers in the context of CPS. Secondly, we introduce a falsification framework, named FalsifAI, that exploits the coverage information for better falsification guidance. Namely, inputs of the controller that increase the coverage (so improving the exploration of the DNN behaviors), are prioritized in the exploitation phase of robustness minimization.

List of Participants

- Aldeida Aleti, Monash University
- Shaukat Ali, Simula Research Laboratory and Oslo Metropolitan University
- Paolo Arcaini, National Institute of Informatics
- Aitor Arrieta, Mondragon University
- Wesley Assunção, North Carolina State University
- Domenico Bianculli, University of Luxembourg
- Christian Birchler, Zurich University of Applied Sciences
- Birthe Böhm, Siemens AG
- Antonio Cicchetti, Mälardalen University
- Myra Cohen, Iowa State University
- Xavier Devroey, University in Namur
- Mattia Fazzini, University of Minnesota
- Alessio Gambi, IMC University of Applied Sciences Krems
- Stefan Klikovits, Johannes Kepler University Linz
- Thomas Laurent, JSPS fellow
- Lei Ma, The University of Tokyo / University of Alberta
- Shane McIntosh, University of Waterloo
- John-Paul Ore, North Carolina State University
- Annibale Panichella, Delft University of Technology
- Sebastiano Panichella, Zurich University of Applied Sciences
- Fabrizio Pastore, University of Luxembourg
- Violet Ka I Pun, Western Norway University of Applied Sciences
- Vincenzo Riccio, Università di Udine
- Volker Stolz, Western Norway University of Applied Sciences
- Catia Trubiani, Gran Sasso Science Institute
- Hironori Washizaki, Waseda University
- Xiao-Yi Zhang, University of Science and Technology Beijing
- Zhenya Zhang, Kyushu University

Figure 2: Overview of the participants

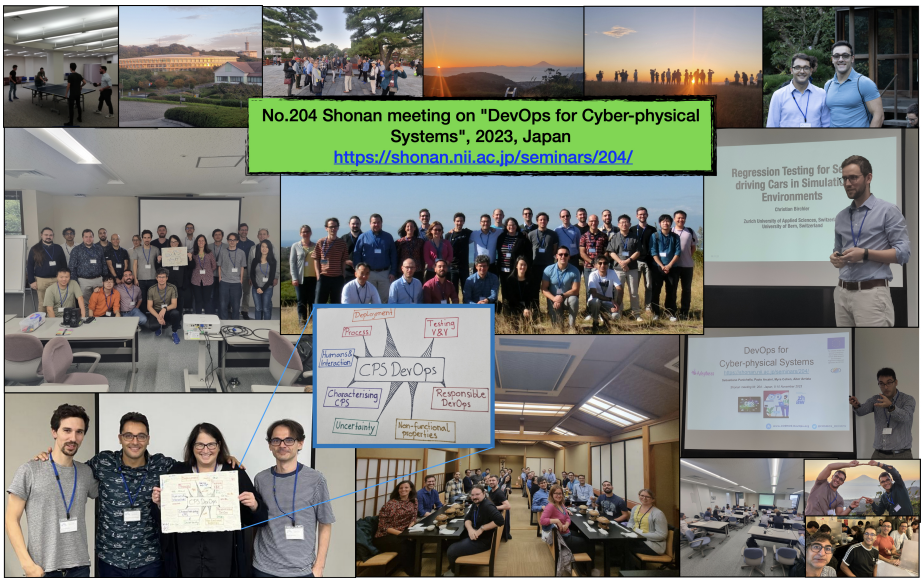

Figure 3: Overview of the meeting outcome

Summary of discussions

The core part of ongoing DevOps solutions requires advanced development and deployment strategies, which work efficiently and in an adaptive manner in different. This Shonan meeting allowed participants to discuss challenges and ideas in all aspects and phases of DevOps in the context of CPSs, including how DevOps can further contribute to improving the efficiency of tomorrow's CPS development and evolution. Specifically, the meeting started with an introductory presentation of the organizers to set the expectations of the meeting. Then, each participant gave a talk on their research related to the meeting's topics, serving as a basis for discussion for the rest of the meeting.

Such talks allowed organizers and participants to identify relevant topics in the context of DevOps for CPSs, which were then discussed in detail within breakout groups (i.e., subsets of participants). The breakout group consisted of structured and open discussion sessions to exchange ideas concerning the following topics:

- Overall CPS: topics such as Characterization (definition, modelling), Case study, Open sourcity, Domains, etc.
- Responsible DevOps: topics such as Human-interaction (Developer, End-user, Data scientist, etc.), Sustainability, Bias, Ethics, Fairness, etc.
- Uncertainty: topics such as Source of uncertainty, etc.
- Environment and Configuration: topic such as Misconfiguration, etc.
- Testing: topics such as Test generation, Regression testing, etc.
- Deployment: topics such as Digital Twins (Reality gap), Runtime monitoring, etc.

- Process: topics such as Code management systems, Configuration management systems, etc.

As discussed in the next section, all these discussions allowed us to identify relevant elements for the future actions to make as next steps in such important research area.

Outcome of the Meeting and the Path Forward

With this Shonan Meeting, we focused on several follow-up activities. Hence, the concrete outcome of the meeting consists of a roadmap of follow-up activities, including the creation of educational material (i.e., a book on DevOps for CPSs), the organization of future international events and summer schools, and the writing of position papers and project proposals. More concretely:

- Book: We already planned upfront to focus on generating the first book on DevOps for CPSs. This book will collect and integrate challenges, bad/best practices, experiences, tools, gaps, and future directions in the field identified during the meeting. The will be written in a coordinated way with participants and organizers of the Shonan meeting, and we expect it will guide future generations of experts moving forward on DevOps and testing automation for complex CPSs. We discussed the main topics during the breakout sessions of the meeting. These topics will be used to design the outline of the book, with the idea to evolve it after the meeting together with all contributors of chapters.

- Organization of Future New/Existing Events: The outcome of the meeting will consist of several elements including the creation of workshops, CPS related competitions (e.g., within ICRA and SBFT communities) [1], event within a conference (e.g., ICST), dedicated special issues in top journals. Complementary, adding related research topics in conferences is planned (e.g., in the CFP).

- Summer schools and other dissemination actions: GSSI (Gran Sasso Science Institute) is committed to hosting PhD schools (some from other universities that joined the event) and supporting/organizing these types of events. Research/Staff mobility actions, including Erasmus + KA 171 and similar frameworks allowing EU and Non-EU visits are part of our dissemination agenda.

- Writing of Position Papers: A number of position papers will be led by participants of the event concerning topics such as testing within DevOps for CPSs uncertainty-aware development for CPSs, and MLOps for CPSs.

- Project proposals: Finally, a list of project proposals planned, which will involve participants from all over the world, to make concrete long-term collaboration happen in a structured way. As concrete example, Sebastiano Panichella, Aitor Arrieta, and Shaukat Ali, who participated in the meeting, are already in the process of finalizing am EU proposal concerning advanced development/testing topics for autonomous CPSs.

[1] As already started for 2024: `https://sbft24.github.io/tools/` and `https://github.com/skhatiri/UAV-Testing-Competition`

From a broader point of view, we believe this meeting will provide a forum to build international cooperation and joint international projects and publications in the future between different communities.